

基于社会特征周期演化的机会移动网络路由转发策略

黄永锋¹, 董永强^{1,2}, 张三峰^{1,2}, 吴国新^{1,2}

(1.东南大学 计算机科学与工程学院, 江苏 南京 211189; 2.东南大学 计算机网络和信息集成教育部重点实验室, 江苏 南京 211189)

摘要: 针对分布式 k 团社区检测引起的超大社区问题, 提出了具有节点退出机制的 τ -window 社区检测方法, 相应提出了 τ -window 中心性估计。通过实验发现 τ -window 社区和 τ -window 中心性具有周期演化特性, 利用该特性, 提出 TTL(time to live)社区检测和 TTL 中心性估计, 以更准确预测消息生存期上节点的相遇。随后, 利用 TTL 社区和 TTL 中心性作为转发测度, 设计了新的机会移动网络路由算法 PerEvo。实验结果表明, 与现有的基于社会特征的路由算法比较, PerEvo 在保持基本不变的传输开销的同时, 有效提高了机会移动网络消息投递的成功率。

关键词: 机会移动网络; 社区; 中心性; 周期演化; 消息转发

中图分类号: TP393

文献标识码: A

Message forwarding based on periodically evolving social characteristics in opportunistic mobile networks

HUANG Yong-feng¹, DONG Yong-qiang^{1,2}, ZHANG San-feng^{1,2}, WU Guo-xin^{1,2}

(1.School of Computer Science and Engineering, Southeast University, Nanjing 211189, China;

2.Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, Nanjing 211189, China)

Abstract: To avoid monster community problem which suffered by distributed k -clique community detection, τ -window community detection was proposed. In addition, τ -window centrality estimation was put forward. By investigating the periodic evolution of τ -window community and τ -window centrality, two new metrics, TTL(time to live) community and TTL centrality, were proposed to improve the prediction of the node's encounter during the message's lifetime. Moreover, a social-aware routing algorithm, PerEvo, was then designed based on them. Extensive trace-driven simulation results show that PerEvo achieves higher message delivery ratio than the existing social-based forwarding schemes, while keeping similar routing overhead.

Key words: opportunistic mobile networks; community; centrality; periodic evolution; message forwarding

1 引言

随着大量短距离、低成本智能设备的普及, 无线自组网应用获得了蓬勃发展。在这类缺乏基础设施支持的网络里, 设备间可能并不存在端到端实时连通路径, 然而, 这并不意味着源和目标间不能实现通信^[1,2]。由于节点的移动, 2 个节点可以进入相互通信范围而

交换数据, 机会移动网络^[3]就是利用节点的机会相遇逐跳将数据从源节点传输到目标的一类网络。

机会移动网络是容迟网络(DTN^[4], delay tolerant network)的一种形式, 其路由转发策略直接影响着数据传输的效率。为了降低传输开销, 往往不采用泛洪(epidemic)策略, 而是把消息仅转发给更有机会接近目标的节点。为此, 研究人员提出了基于相

收稿日期: 2013-09-28; 修回日期: 2014-05-07

基金项目: 国家自然科学基金资助项目(61272532, 61370209, 61300200, 61472080); 国家高技术研究发展计划(“863”计划)基金资助项目(2013AA013503); 江苏省博士后科研基金资助项目(1402039B); 江苏省未来网络前瞻性研究基金资助项目(BY2013095-2-06)

Foundation Items: The National Natural Science Foundation of China(61272532, 61370209, 61300200, 61472080); The National High Technology Research and Development Program (863 Program) of China(2013AA013503); Jiangsu Planned Project for Postdoctoral Research Funds(1402039B); The Future Networks Prospective Research Program of Jiangsu Province(BY2013095-2-06)

遇概率进行效用计算的转发策略, 如 Prophet^[5]和 Maxprop^[6]。然而, 这些策略基于历史直接计算到其他节点的相遇概率, 预测的稳定性因节点移动的时变性而受到影响, 因此, 近年来, 利用节点更稳定的社会特征进行相遇预测引起了研究者的极大兴趣^[7, 8]。

社区和中心性是常被使用的社会特征。社区表征节点所属的社会群体, 社区内个体比社区间个体的联系一般更为紧密, 这种紧密程度通常用模度^[9]来表示, 针对某一特定社区划分, 模度 Q 如式(1)所示。

$$Q = \sum_i (e_{ii} - a_i^2) \quad (1)$$

其中, e_{ii} 是社区 i 内实际的边占系统中边的百分比, 而 a_i^2 是随机情况下社区 i 内的边占系统中边的百分比, Q 越大, 社区内部联系一般越紧密。中心性表征了节点的联系能力, Freeman 等人在文献[10]中提出了 3 种最广泛使用的中心性, 其中, 节点转发消息的能力被描述为介数中心性, 如式(2)所示。

$$C_B(v_k) = \sum_i \sum_{j < k} \frac{g_{ij}(v_k)}{g_{ij}} \quad \text{s.t. } i \neq j \neq k \quad (2)$$

其中, g_{ij} 表示连接节点 v_i 和 v_j 的最短路径数量, 而 $g_{ij}(v_k)$ 表示这些最短路径中经过 v_k 的路径数量, 由式(2)可知, 介数中心性体现了节点在其他节点通信过程中的作用。

根据文献[7, 11], 单纯地利用社区和中心性的转发策略在某些情况下效果并不理想, 而社会特征路由 Bubble Rap (简称 Bubble) 巧妙地将中心性与社区结合起来, Bubble 在传输消息时, 首先向全局范围内中心性更高的节点进行转发, 直到消息到达目标社区, 接着向社区范围内中心性更高的节点进行转发, 直到消息到达目标节点, 其详细转发过程如图 1 所示。文献[7]表明, Bubble 在转发性能上要优于基于相遇概率的转发策略 Prophet。

虽然利用社区和中心性作为转发测度增强了路由转发的稳定性, 但 Bubble 并没有充分考虑路由的动态响应性, 在社区检测和中心性估计方面存在明显不足: 一方面, Bubble 使用的分布式社区检测方法, 没有提供节点及时退出社区的机制, 随着时间演化, 系统可能蜕变成少数超大社区, 这种社区划分无法再改善消息的转发性能; 另一方面, 由于被路由的消息大多具有时效性, 现有的社区划分和中心性估计方法, 基于整个历史的累积联系, 没有

考虑不同时段节点社会联系的异质性, 导致 Bubble 的社区划分和中心性估计方法, 无法准确反映节点在消息生存期内的联系行为。

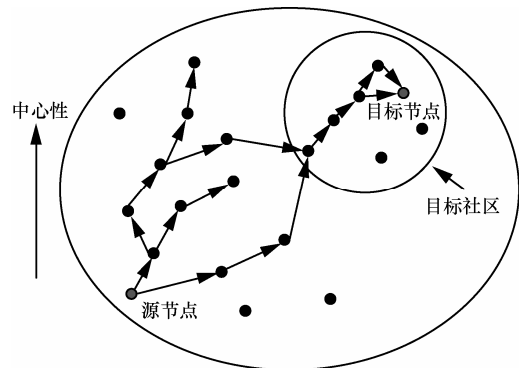


图 1 Bubble Rap 路由转发策略

文献[12]显示, 93%的移动行为是可预见的, 同时, 文献[13, 14]表明, 机会移动网络中人们的交互具有一定的周期性, 人们可能每天在相同的时间段和相同的人相遇, 因此, 考虑节点的社会特征是否也有这样的周期性, 如果是, 可否利用与消息生存期相同历史时间段而非整个历史的社会特征来预测节点行为, 改进当前的路由转发策略。

本文的主要贡献包括: 首先, 引入节点退出社区的机制, 弥补了 Bubble 所使用的社区检测方法引起的超大社区问题, 提出了基于时间窗的 τ -window 社区检测和 τ -window 中心性估计方法; 其次, 利用 τ -window 社区检测和 τ -window 中心性估计对实证数据集分析, 发现了节点的社会特征存在周期演化特性, 即不同周期相同时间窗、同一周期相邻时间窗的社会特征相似(相关), 而同一周期非相邻时间窗的社会特征存在明显差异; 再次, 基于社会特征周期演化特性, 提出 TTL(time to live)社区检测和 TTL 中心性估计方法以动态预测消息生存期内的社区演化和节点的中继能力, 对节点的相遇概率进行更加准确的估计。最后利用 TTL 社区检测和 TTL 中心性估计方法, 设计了一种新的机会移动网络路由转发策略 PerEvo, 该策略充分考虑了节点社会特征周期演化的特性, 仿真实验表明, PerEvo 策略在路由性能上优于现有的基于社会特征的路由机制。

2 相关工作

Simbet^[8]是机会移动网络中较早把社会特征作为效用计算的路由策略, 该策略混合了社区和中心性 2 种社会特征, 虽然没有明确使用社区概念, 但

在效用计算时，通过考虑与目标节点公共邻居的数量隐性地比较了与目标节点的社会关系。由于介数中心性^[15]被认为能够较好地体现节点在消息转发过程中的地位，且个体网络^[16](ego network)中的介数中心性与节点在全局网络中的介数中心性在大小次序上存在较强的相关性，容易分布式实现，因此，效用计算中节点的中心性采用个体网络介数中心性进行度量。实验表明，Simbet 较传统的非社会特征路由有更好的性能。

相比 Simbet, Bubble 区分对待了社区和中心性，实验表明，其性能要略优于 Simbet。Bubble 的中心性理论上采用了介数中心性，为便于实现，实际中，Bubble 选择了历史平均单位时间联系其他节点的数量作为中心性估计，并且实验论证了两者的相关性，但这种基于整个历史平均的中心性估计，无法体现不同时段节点中继能力的差异。在社区检测方法上，Bubble 采用的方法是基于 k 团的分布式社区检测^[17]，该方法源于集中式 k 团检测^[18]，能够揭示现实中的重叠社区结构，这种结构被认为可提高消息转发性能^[19]，然而，由于节点缺乏退出社区的机制，系统可能逐步蜕变成少数超大社区，对数据转发而言，这种超大社区的划分没有意义。

为了避免超大社区问题，文献[20]提出了一种基于期望的分布式时空社区检测方法 DEBT, DEBT 在每个离散时间窗的尾部计算所有相遇节点在最近 w 个窗口上与当前节点的平均窗口联系时间，并以此作为社区成员的接纳基准。虽然 DEBT 方法可及时反映社区成员的动态变化，但由于其本质上是一个以被检测节点为中心的社区检测方法，社区成员仅限于发生直接联系的节点，在改进消息转发性能上也很难再与中心性相结合。

文献[21]是一种同时利用社区和中心性作为转发测度的多播路由策略，其基于累积联系频率检测社区，基于累积联系概率计算节点中心性，无法准确反映节点当前时间段下的社区划分和联系能力，因此，文献[22]提出瞬态社会联系模式，在大量实证数据基础上，通过把瞬态联系分布、瞬态连接和瞬态社区 3 种联系模式统一成高斯分布形式，提出了新的中心性设计，改善了 DTN 路由性能。

本文路由算法 PerEvo 在基本思路与 Bubble 一脉相承，避免了文献[20]提到的蜕变成超大社区的问题，同时借鉴了文献[22]关于瞬态社会联系的思想。本文的创新在于，发现了机会移动网络节点

的社会特征存在周期演化特性，利用节点在历史周期与消息生存期相同时间段上的联系行为，改进了社区和中心性的度量，无论文献[20]还是文献[22]，均没有考虑社会特征周期演化特性，提出的社区检测和中心性估计方法也无法体现节点在消息生存期内的社会联系。

为了区分节点在不同时段的社会行为，本文首先对人类的社会活动按时间窗进行划分，提出了基于时间窗的 τ -window 社区检测和中心性估计方法，并以此为基础挖掘人类活动的行为规律。

3 基于时间窗的社区检测与中心性估计

3.1 实验数据集

本章及后文的实验基于 MIT Reality 数据集^[13]，该数据集包含对 97 个智能手机用户长达 9 个月跟踪的数据，数据项记录了带有蓝牙功能的手机之间的相遇信息，手机用户通过 ID 唯一标识。

3.2 τ -window 社区检测

本节介绍的 τ -window 社区检测源于文献[17]的分布式 k 团社区检测方法，假设当前节点为 v_0 ，相遇节点为 v_i ，下面对文献[17]中的定义和社区检测方法做相关介绍。

定义 1 熟人集： v_0 的熟人集是与 v_0 的累积联系时间超过阈值 F_{thresh} 的节点集，表示为 F_0 ， v_i 的熟人集表示为 F_i 。对于节点 v_i ，它完全了解自己的熟人集 F_i ，但由于 v_0 了解的信息与 v_i 并不完全同步， v_0 了解的 F_i 称为 F_i 在 v_0 节点上的近似，表示为 \tilde{F}_i 。

定义 2 本地社区： v_0 的本地社区表示为 C_0 ，包含 v_0 熟人集中的所有节点以及由分布式 k 团社区检测算法选择的其他节点， v_i 的本地社区表示为 C_i 。

对于 v_0 ，分布式 k 团社区检测算法就是要实时维护节点的本地社区 C_0 ，为此， v_0 需要维护下列信息：与其他节点的相遇历史，熟人集 F_0 ， C_0 内节点的熟人集在 v_0 上近似的集合 $FSoLC_0$ 。

$$FSoLC_0 = \{\tilde{F}_j \mid v_j \in C_0\} \quad (3)$$

设初始时 $C_0 = \{v_0\}$ ， $F_0 = \emptyset$ ， $FSoLC_0 = \emptyset$ ，当 v_0 遇到 v_i 时， v_0 执行以下算法步骤。

步骤 1 v_0 从 v_i 处获得 C_i ， F_i 和 $FSoLC_i$ ， v_0 首先更新关于 v_i 的熟人集的近似 $\tilde{F}_i = F_i$ ，然后根据获得的 $FSoLC_i$ 更新 $FSoLC_0$ ，即利用 $FSoLC_i$ 中时间更新的熟人集的近似替代 $FSoLC_0$ 中相应的熟人集的近似。

步骤2 假如 v_i 不属于 F_0 , 且直到与 v_0 失去联系时两者的累积联系时间仍未超过阈值 F_{thresh} , 则执行步骤3, 否则一旦超过指定阈值 F_{thresh} , v_0 便将 v_i 加入 F_0 和 C_0 , 设置 $\tilde{F}_i = F_i$ 并加入 $FSoLC_0$, 然后执行步骤4。

步骤3 假如 v_i 的熟人集 F_i 至少包含 C_0 中的 $k-1$ 个节点, 即满足

$$|F_i \cap C_0| \geq k-1 \quad (4)$$

时, 把 v_i 加入 C_0 , 设置 $\tilde{F}_i = F_i$ 并加入 $FSoLC_0$ 。

步骤4 假如 v_i 已加入 C_0 , 则考虑 C_i 中的节点 (设为 v_j) 是否可加入 C_0 , 如果 v_j 不在 C_0 里且 v_j 的熟人集在 v_0 上的近似 \tilde{F}_j 至少包含 C_0 中的 $k-1$ 个节点, 即满足

$$|\tilde{F}_j \cap C_0| \geq k-1 \quad (5)$$

时, v_j 就加入 C_0 , 相应地设置 $\tilde{F}_j = F_j$, 并加入 $FSoLC_0$ 。

以上算法实时维护的本地社区 C_0 就是 v_0 的社区, 该算法在线检测时存在一个问题, 即随着时间演化, C_0 将蜕变成超大社区。图2显示了该算法在线检测时某节点每天 12:00 的社区尺寸, 由图2可知, 由于节点缺乏退出社区机制, 大部分节点最后位于同一社区, 这对于消息转发已经没有意义。

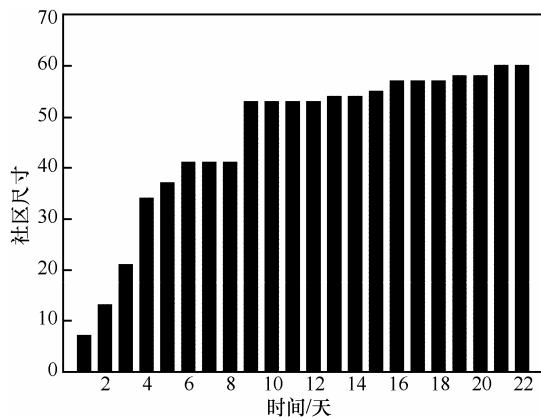


图2 分布式 k 团算法检测的某节点社区的尺寸变化(ID 为 29)

为了避免以上问题, 本文按人类活动规律, 把原有节点联系划分成连续的时间窗(时长为 τ , 也称 τ -window), 在进行与指定阈值 F_{thresh} 的比较时, 使用新的加权累积联系时间 T^w 替代原有的累积联系。

定义3 加权累积联系时间 T^w : T^w 定义如式(6)所示, T_t^w 表示第 t 个时间窗节点间加权累积联系时间, T_t^c 为第 t 个时间窗节点间的累积联系时长。

$$\begin{cases} T_1^w = T_1^c \\ T_t^w = \alpha T_{t-1}^w + (1-\alpha)T_t^c \quad (t > 1) \end{cases} \quad (6)$$

其中, α 为老化指数, 范围为 $[0,1]$ 。

改进的社区检测方法以分布式 k 团检测为基础, 在每个时间窗的尾部(也是新时间窗的起始)增加计算加权累积联系 T^w , 并引入节点退出社区的机制, 由于时间窗长度为 τ , 改进的方法称为 τ -window 社区检测, 在各时间窗尾部增加的执行过程如下。

步骤5 初始化 T_t^c 为 0, 根据式(6)计算加权累积联系 T^w , 对 F_0 、 C_0 以及 $FSoLC_0$ 进行更新, 从集合中移除长时间没有联系的节点, 更新过程如下。

步骤6 更新 F_0 。由于 T_t^w 重新计算后可能变小, 从 F_0 中移除 T_t^w 低于 F_{thresh} 的节点生成新的 F_0 。原有 C_0 现分成 2 部分, 一部分是新 F_0 中的节点, 一部分是不在新 F_0 中的节点, 设为 G_0 , 如式(7)所示。

$$G_0 = \{v_k \mid v_k \in C_0 \cap v_k \notin F_0\} \quad (7)$$

同理, 对于 C_0 内任意节点 v_j 的熟人集在 C_0 处的近似 \tilde{F}_j 也按照 F_0 的更新方式进行更新(注: v_0 上需保存计算 v_j 与其熟人加权累积联系时间相关的近似信息)。

步骤7 更新 C_0 。令 $C_0 = F_0$, 不断从 G_0 中挑选符合式(5)的节点移入 C_0 , 直到 G_0 中无法挑出这样的节点, 生成新的 C_0 。

步骤8 更新 $FSoLC_0$ 。新的 C_0 生成后, 根据式(3)生成新的 $FSoLC_0$ 。

通过上述方法检测时, 节点在 t 时刻维护的本地社区 C_0 定义为 t 时刻节点的 τ -window 社区, 当 α 较小时, τ -window 社区更多反映节点最近的社会联系, 由于引入了节点退出社区机制, C_0 不会再变成超大社区, 这点在社区的周期演化分析中得到了验证。

3.3 τ -window 中心性估计

Bubble 实际选择的 c -window 中心性度量基于节点在单位时间内联系不同节点的数量, 沿用该方法并与 τ -window 社区相对应, 节点最近 τ 时长内联系的不同节点数量定义为节点的 τ -window 中心性, τ -window 中心性反映了节点最近的联系能力。

4 社会特征周期演化特性分析

τ -window 社区和 τ -window 中心性更侧重反映节点的最近联系, τ -window 社区检测避免了原有分布式 k 团的一些问题, 但还不足以描述节点在消息

生存期上的社会行为， τ -window 中心性也不足以描述节点在消息生存期上的中继能力，需要基于这 2 种度量进一步观察节点在消息生存期上的联系。

本章基于 τ -window 社区和 τ -window 中心性，对机会移动网络的社会特征演化进行分析，由于数据转发效率取决于活跃节点行为，随机抽样了数据集^[13]5 个活跃节点 1 个月内 22 个连续工作日的数据观察分析，以天为周期，时间窗长度 τ 取 3 个小时，主要观察时间窗尾部 τ -window 社区和中心性的演化，为简化表述，这样的社区和中心性分别称为时间窗上的 τ -window 社区和 τ -window 中心性。

4.1 τ -window 社区的周期演化

本节考察时间窗上的 τ -window 社区随时间演化的特征，重点挖掘 τ -window 社区的演化规则，包括社区尺寸变化规律，社区相似程度。分析中加权累积联系的老化指数 α 取 0.1，阈值 F_{thresh} 取 300 s。

图 3(a)展示了 ID 为 29 的节点连续 5 个工作日的社区尺寸变化，图 3(b)展示了抽样节点平均的社区尺寸变化，由图 3 可知，单个节点和系统整体的 τ -window 社区尺寸变化均展现出较强的周期行为。

设节点在不同时间窗上的 τ -window 社区分别为 C_1 和 C_2 ，为了衡量 2 个社区的相似程度，利用 Jaccard 相似度量定义 C_1 和 C_2 的社区相似度量如式(8)所示。

$$J(C_1, C_2) = \frac{|C_1 \cap C_2|}{|C_1 \cup C_2|} \quad (8)$$

该定义通过 2 个社区集合中公共节点的数目占总节点数目的比值来表征社区的相似程度，根据该定义容易计算，大小相同的具有 70%公共节点的 2 个社区之间的社区相似度量约为 0.5。

假设 J_k^i 为节点 v_i 在 2 个周期第 k 个时间窗上 τ -window 社区的相似度量， ω 为一个周期时间窗的数目，则节点 v_i 连续周期的社区相似度量 S_τ^i 定义为两周期中各相同时间窗 τ -window 社区相似度量的平均。

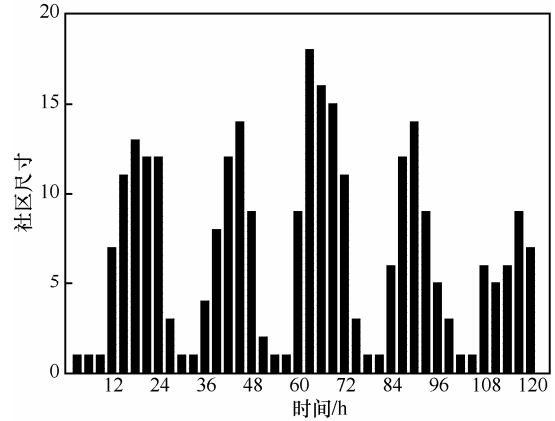
$$S_\tau^i = \left(\sum_{k=1}^{\omega} J_k^i \right) / \omega \quad (9)$$

设抽样节点数为 q ，抽样数据集连续周期的社区相似度量 S_τ 表示如下

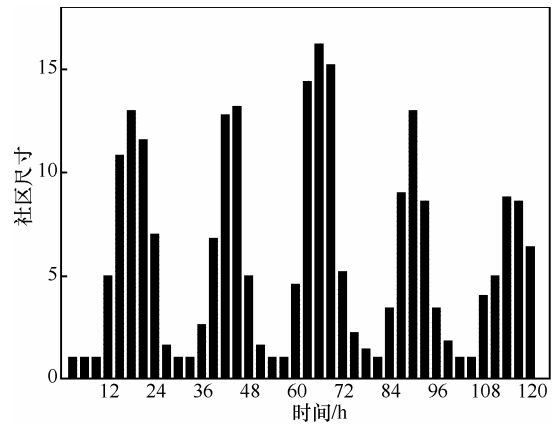
$$S_\tau = \left(\sum_{i=1}^q S_\tau^i \right) / q \quad (10)$$

图 4 展示了 S_τ 的变化，水平虚线表示均值为 0.51，从式(8)及图 4 可知，当前周期某时间窗节点

v_i 密切联系的节点中，约 70%在下一周期相同时间段仍将与 v_i 保持密切联系。



(a) ID 为 29 的节点的社区尺寸



(b) 活跃节点平均社区尺寸

图 3 活跃节点连续 5 天的社区尺寸变化

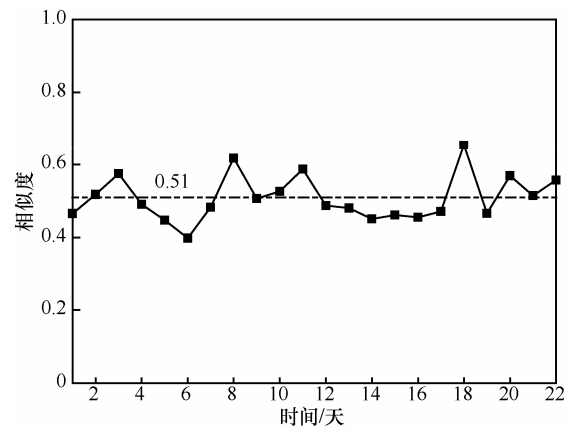


图 4 活跃节点连续工作日的社区相似度量

与 S_τ^i 和 S_τ 定义类似， W_τ^i 定义为节点 v_i 一个周期中相邻时间窗 τ -window 社区相似度量的平均， W_τ 为所有抽样节点 W_τ^i 值的平均，图 5 展示了 W_τ 的变化。由图 5 可知，节点的活动表现出一定的连续性，

即前一时间窗属于 v_i 社区内的节点, 仍然有一部分在接下来的时间窗继续逗留在 v_i 社区内。

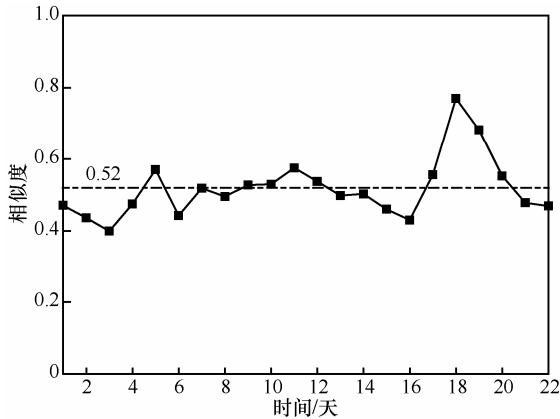


图5 活跃节点相邻时间窗的社区相似度

同理, 节点非相邻时间窗社区相似度(仅考虑间隔时间窗) P_{τ}^i 定义为节点 v_i 一个周期中非相邻时间窗 τ -window 社区相似度的平均, P_{τ} 为所有抽样节点 P_{τ}^i 值的平均, 图6展示了 P_{τ} 的变化, 由图6可知, 节点一个周期内非相邻时间窗社区差异较大。

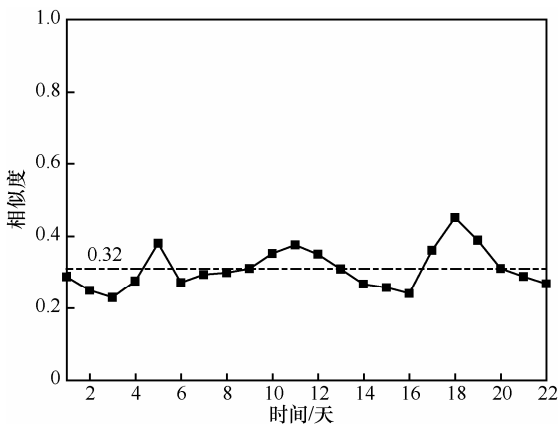


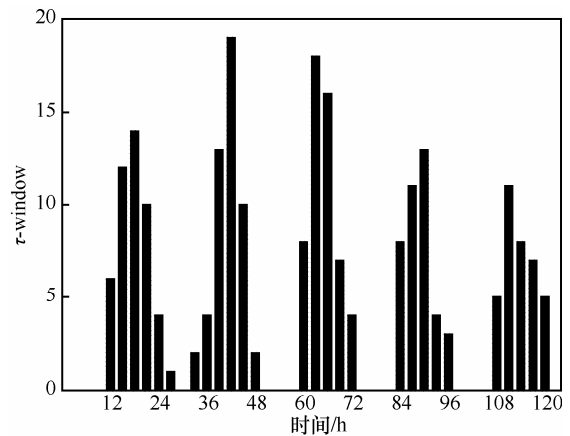
图6 活跃节点非相邻时间窗的社区相似度

以上实验分析表明, 节点的 τ -window 社区按照规律周期性变化, 连续周期相同时间窗及同一周期相邻时间窗社区相似度较高, 而同一周期非相邻时间窗的社区存在明显差异, 即节点的社区存在周期演化特性。

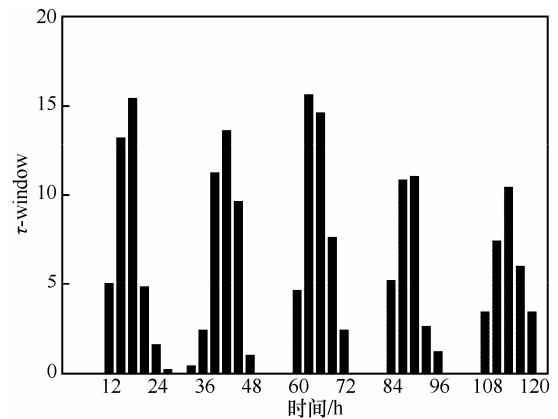
4.2 τ -window 中心性的周期演化

本章重点考察 τ -window 中心性的变化规律及不同时间窗 τ -window 中心性的相关性。图7为抽样节点连续5个工作日的 τ -window 中心性的变化, 部分时间窗的 τ -window 中心性由于为0而没有显示, 由图7可知, 单个节点和系统整体的 τ -window 值变

化均展现出较强的周期行为。实验还发现, 连续周期相同时间窗节点的 τ -window 中心性皮尔逊相关系数为 0.81, 两者存在较强相关性; 同一周期相邻时间窗节点的 τ -window 中心性相关系数为 0.54, 两者存在一定相关性; 而同一周期非相邻时间窗(此处仅考虑间隔时间窗) 节点的 τ -window 中心性相关系数为 -0.02, 两者不相关。



(a) ID 为 29 的节点的 τ -window 中心性



(b) 活跃节点平均 τ -window 中心性

图7 活跃节点连续5天的 τ -window 中心性变化

以上分析表明, 节点 τ -window 中心性也存在周期演化特性, 不同周期相同时间窗 τ -window 中心性及同一周期相邻时间窗的 τ -window 中心性相关, 而同一周期非相邻时间窗的 τ -window 中心性不相关。

4.3 周期演化特性的总结

通过 τ -window 社区和 τ -window 中心性的实验分析发现, 节点社会特征的演化展现了较强的规律性, 这符合人们日常社会活动的一般特征。节点在相邻时间窗上社会联系的相似性或相关性, 与人类社会活动的延续性相吻合, 而节点在同一周期非相邻时间窗上社会联系的异质性, 表明社会活动中个

体在不同时间段的联系行为存在很大差异，也意味着以往单纯地基于累积联系划分社区和进行中心性估计的方法存在较大缺陷，无法体现这种差异。

5 基于消息生存期的社区检测和中心性估计

5.1 社区和中心性的讨论

由于被路由的消息大多具有时效性，对于以社区和中心性为转发测度的路由策略，要提高其路由性能，关键要让节点的社区和中心性度量能够准确预测节点在消息生存期内的联系行为。

对于分布式 k 团算法，文献[17]表明，针对后验分析，如果消息生存期为 $[T, T_e]$ ，根据 $(T_e - T)$ 时长设置相应阈值 F_{thresh} ，在 T 时刻启动算法，该算法在 T_e 时刻检测的社区 C_0 可准确反映 $[T, T_e]$ 时段上 v_0 的社会联系。然而，实际路由消息时， T 时刻就要预测节点的社区，此时节点并不知道 $[T, T_e]$ 时段的联系，而且由于算法面向节点而非消息运行，无法依据单个消息在 T 时刻启动算法，也无法依据单个消息的生存期设置 F_{thresh} ，因此，利用该算法在 T 时刻得到未来消息生存期上 v_0 的社区并不可行。该算法实际在线检测时产生的问题在 3.2 节已详细介绍。

对于 τ -window 社区和中心性，其侧重反映节点最近的行为，由于节点在不同时段的联系行为存在很大差异， τ -window 社区和中心性也不足以表征节点在消息生存期上的联系。

就 τ -window 社区来说， α 取值较小时，社区反映的是与节点最近 τ 的社会联系，由于节点在非相邻时间窗的社区展现了较强的异质性，未来的联系可能与当前有很大不同， τ -window 社区可能会忽略好的潜在中继。 α 取值较大时， τ -window 社区就是一种累积意义上的社区，无法反映节点在不同时间窗联系的异质性，可能引起更大的转发开销，例如， v_i 可能经常与 v_j 在上午联系，而与 v_m 在晚上联系，对于生存期仅有 6 h 的消息，要投递给 v_i ，上午转发给 v_j ，晚上转发给 v_m 更为合适，然而， α 较大时， v_j 和 v_m 可能都是 v_i 的 τ -window 社区成员。

对于 τ -window 中心性，同样由于未来联系可能与当前有很大不同，无法准确预测好的潜在中继。而像 Bubble 中的 c -window，等同于所有历史时间窗平均的 τ -window 中心性，也无法反映节点在不同时间窗联系能力的异质性。图 8 显示了 3 个节点在不同时间窗的 τ -window 和 c -window，分别记为 λ_i

和 λ_c (τ 取 6)，如果 v_0 在 6:00 遇到 v_i ，消息过期时间是 12:00，或者在 12:00 遇到 v_j ，过期时间是 18:00，根据 τ -window 和 c -window，消息都不应转发，然而，由于相遇节点在消息过期前展现了比当前节点 v_0 更强的联系能力，转发将提高投递的成功率。

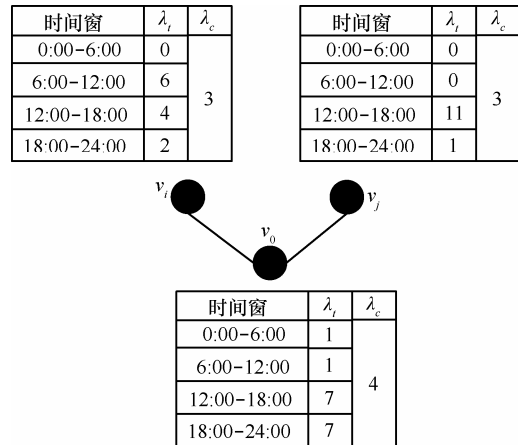


图 8 τ -window 与 c -window 中心性

虽然 τ -window 社区和中心性不足以预测节点在整个消息生存期上的联系，但它们的演化展现了较强的规律性，本文利用这种规律性，基于 τ -window 社区和中心性，提出能够更准确预测节点消息生存期内联系行为的 TTL 社区和 TTL 中心性。

5.2 TTL 社区生成

定义 4 TTL 社区：表示为 $Tcom(v_0, m)$ ，是对节点在消息生存期内社会联系的一种预测，如果目标节点在 v_0 针对 m 的 TTL 社区内，表明 v_0 在消息过期前有较大的概率与目标联系。

设 T 为消息 m 被转发的时间， T_e 为消息过期时间， v_0 关于 m 的 TTL 社区基于以下社区生成，一是节点 T 时刻的 τ -window 社区 C_T ；二是节点历史周期中与消息 m 生存期相同时间窗的 τ -window 社区数组 C_{11} 、 C_{12} 、 \dots 、 C_{1K} 、 C_{21} 、 C_{22} 、 \dots 、 C_{2K} 、 \dots 、 C_{L1} 、 C_{L2} 、 \dots 、 C_{LK} ，如图 9 所示。显然，为了生成 TTL 社区；系统需保存节点最近 L (预设参数) 个周期时间窗上的 τ -window 社区。

详细的 TTL 社区生成过程如下：首先， C_T 中的成员反映了节点最近的联系，这些成员全部加入 TTL 社区。其次，为了控制开销，如果节点在 τ -window 社区数组中的出现次数超过 CC_{thresh} (预设阈值)，该节点也加入 TTL 社区。一般来说，消息生存期越长，历史周期数越大， CC_{thresh} 也越大，此处 CC_{thresh} 取 λKL ， K 为消息生存期跨越的时间窗数，

L 为预设周期数, λ 为预设常数。详细的 TTL 社区生成过程如算法 1, 相关符号如表 1 所示。

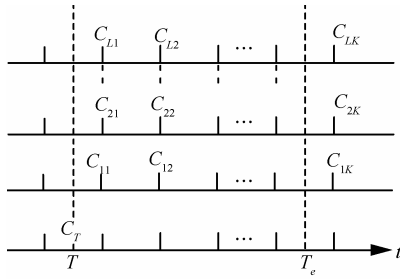


图9 TTL 社区生成

表 1 各种函数及符号含义

函数	含义
Tcom	针对消息 m 的 TTL 社区
C_T	节点 T 时刻的 τ -window 社区
getTtl(m)	消息 m 的生存时间
λ	TTL 社区准入参数
K	消息 m 生存期跨越的时间窗数
L	预设周期数
CC_{thresh}	消息生存期 TTL 社区判定阈值
τ	时间窗长度
ArrayC	图 9 的 τ -window 社区数组
inCount(v_j , ArrayC)	v_j 在 ArrayC 中出现的次数

算法 1 Tcom(v_0, m)

- 1) for each node $v_j \in C_T$ do
- 2) tcom.add(v_j);
- 3) end for
- 4) msgttl = getTtl(m);
- 5) $K = \text{msgttl} / \tau$;
- 6) ArrayC = getTauWindowCommunity(K, L, m);
- 7) for each node v_j in system do
- 8) if inCount(v_j , ArrayC) $> CC_{\text{thresh}}$ then
- 9) tcom.add(v_j)
- 10) end for
- 11) return tcom;

5.3 TTL 社区分析与评估

TTL 社区依赖 T 时刻节点的 τ -window 社区, 在于节点联系活动的延续性及其当前周期联系行为存在有别于历史的随机性; 其依赖历史周期中与消息生存期相同时间窗的 τ -window 社区, 在于节点联系行为的周期性。算法仅包括经常出现在历史周期

时间窗 τ -window 社区里的节点, 只有这些节点, 才与节点在消息生存期上有更稳定的社会联系。

为评估 TTL 社区预测的准确程度, 需观察 TTL 社区内的节点在消息生存期内是否真实与该节点紧密联系。5.1 节提到, 对于后验分析, 依然可使用分布式 k 团来准确反映消息生存期上节点的社会联系, 因此, 在 T 时刻启动该算法, 基于 $[T, T_e]$ 时段实际发生的联系检测节点在消息生存期上的实际社区, 使用 4.1 节式(8)中的 Jaccard 相似度比较 TTL 社区与消息生存期内实际社区的相似程度。进行多轮评估, 每轮评估中, 当前节点 v_0 从活跃节点中随机挑选, 消息生存期起点 T 和终点 T_e 随机生成, 相似度取多轮实验的平均值。图 10 是相似度随社区老化指数 α 的变化, 为形成对比, 图中也显示了 T 时刻 τ -window 社区与实际社区的相似度。

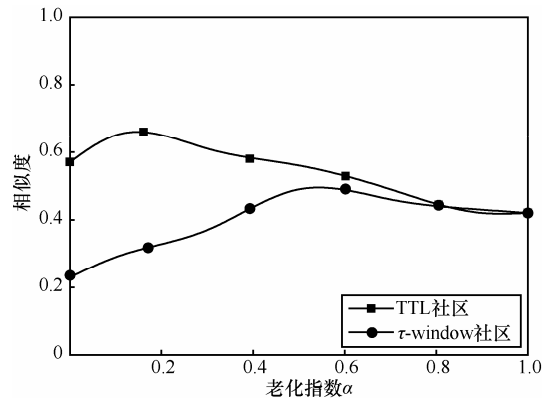


图 10 TTL 社区、 τ -window 社区与实际社区的相似度

由图 10 看出, TTL 社区较 τ -window 社区更能准确预测消息在生存期内的联系行为, 通过图 10, 也可得到老化指数 α 的优化设置约在 0.1 左右。

5.4 TTL 中心性估计

定义 5 TTL 中心性: 表示为 $Tcen(v_0, m)$, 是对节点在消息 m 生存期内联系能力的预测。如果 v_0 的 TTL 中心性大于 v_i , 表明 v_0 在消息生存期内有更强的中继能力, 如果预测范围是整个系统, 该中心性称为全局 TTL 中心性, 表示为 $GTcen(v_0, m)$, 如果预测仅限于针对 m 的 TTL 社区, 该中心性称为社区 TTL 中心性, 表示为 $LTcen(v_0, m)$ 。

与 TTL 社区类似, TTL 中心性通过两方面估计: 一是节点当前联系能力, 使用节点 T 时刻的 τ -window 中心性表示; 二是节点历史周期中与转发消息生存期相同时间段的联系能力, 使用节点在历史周期中跨越消息生存期时间窗的平均 τ -window

中心性表示。设消息 m 的转发时刻为 T ，过期时刻为 T_e ，TTL 中心性依赖关系如图 11 所示。

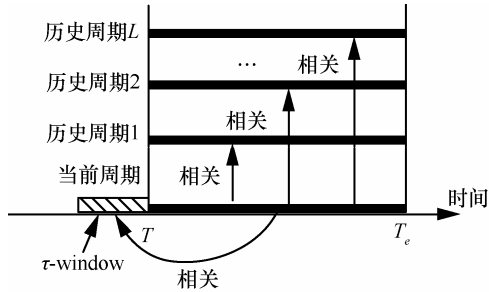


图 11 TTL 中心性估计

本文采取简单加权和的办法处理以上两者关系，权重由 τ -window 时间窗长度和消息生存期时长决定，由此得到 v_0 针对消息 m 的全局 TTL 中心性 $GTcen(v_0, m)$ 如下

$$GTcen(v_0, m) = \frac{\tau}{\tau + (T_e - T)} tc_T + \frac{T_e - T}{\tau + (T_e - T)} tc_L \quad (11)$$

其中， tc_T 为节点 T 时刻的 τ -window 中心性， tc_L 为节点历史周期中 LK 个时间窗的平均 τ -window 中心性。同理， v_0 针对消息 m 的社区 TTL 中心性 $LTcen(v_0, m)$ 可在消息 m 的 TTL 社区范围内计算。

5.5 TTL 中心性分析与评估

为了分析 TTL 中心性对节点中继能力预测的准确程度，需要评估 TTL 中心性与消息中继能力之间的相关性，由于时间介数中心性^[23]能够更准确反映消息生存期内节点的中继能力，评估中使用时间介数中心性表征节点中继消息的能力。评估实验中，评估节点从活跃节点中随机挑选，时间介数中心性采用数值计算方法得到，即挑选多对非评估节点作为消息的源和目标节点，并设置消息的生存期为 $[T, T_e]$ ， T 和 T_e 随机生成，从 T 时刻起在源节点处泛洪消息， T_e 时刻后，按源和目标节点对对成功投递的消息进行分组，对每一分组，计算经过评估节点并以时间最短路径^[23] (temporal shortest path) 投递的消息数与所有以时间最短路径投递的消息数的比值，时间介数中心性就是所有分组比值的和。评估结果显示，节点的 TTL 中心性与时间介数中心性相关系数为 0.87，而 τ -window 与 c -window 中心性与时间介数中心性的相关系数分别为 0.61 和 0.73，因此，TTL 中心性能够更准确反映节点在消息生存期内的联系能力。

6 PerEvo 路由转发策略及算法分析

6.1 PerEvo 路由转发策略

本文的路由转发策略 PerEvo 与 Bubble 一脉相承，与 Bubble 相同的是，PerEvo 在消息到达目标社区前，节点依据全局中心性复制消息，消息到达目标社区后，节点在社区内依据社区中心性复制消息，为了控制开销，节点转发消息到目标社区后不再保留副本。与 Bubble 不同的是，算法中所用的社区和中心性是针对消息的 TTL 社区和 TTL 中心性。

设当前节点为 v_0 ，相遇节点为 v_i ，不失一般性，假设消息可在一次机会联系中传输完成，PerEvo 路由转发策略如算法 2 所示，符号含义如表 2 所示。

算法 2 PerEvo-Routing(v_0, v_i)

- 1) for each message $m \in \text{buffer}(v_0)$ do
- 2) $d = \text{dst}(m)$
- 3) if $v_i = d$ then
- 4) node v_0 forwards m to node v_i
- 5) else if $d \in \text{TCom}(v_i, m)$
- 6) && $d \notin \text{TCom}(v_0, m)$ then
- 7) node v_0 forwards m to node v_i
- 8) else if $d \in \text{TCom}(v_0, m)$ && $d \in \text{TCom}(v_i, m)$
- 9) && $LTcen(v_i, m) > LTcen(v_0, m)$ then
- 10) node v_0 copys m to node v_i
- 11) else if $d \notin \text{TCom}(v_0, m)$ && $d \notin \text{TCom}(v_i, m)$
- 12) && $GTcen(v_i, m) > GTcen(v_0, m)$
- 13) node v_0 copys m to node v_i
- 14) end if
- 15) end for

表 2 算法 2 函数符号的含义

函数	含义
$\text{buffer}(v_0)$	v_0 缓存区中的消息集合
$\text{dst}(m)$	消息 m 的目的节点
$\text{Tcom}(v_0, m)$	v_0 针对 m 的 TTL 社区
$LTcen(v_0, m)$	v_0 针对 m 的社区 TTL 中心性
$GTcen(v_0, m)$	v_0 针对 m 的全局 TTL 中心性

6.2 算法的优劣分析

根据 5.3 节的分析，相比 τ -window 社区，与目标节点位于同一 TTL 社区内的节点，在消息生存期上有更大概率与目标发生联系；根据 5.5 节的分析，TTL 中心性，相比 τ -window 和 c -window，也更体现节点在消息生存期的中继能力。因此，对于具有

时效性的消息，PerEvo 的路由性能理论上优于 Bubble。

在 PerEvo 中，最坏情况下，不同的消息可能会对完全不同的 TTL 社区，但由于算法保存了最近 L 个周期时间窗的 τ -window 社区，TTL 社区的生成并不需要重新进行社区检测，而仅需要执行算法 1 中对已保存的 τ -window 社区的简单统计，增加的计算量可以忽略，因此，计算上并不存在性能瓶颈。

7 实验及结果分析

7.1 环境设置

实验基于机会网络环境 ONE^[24]，使用 MIT Reality 数据集一个月中连续工作日的相遇数据，消息的创建时间随机分布在连续的 5 个工作日内，实验分成 5 组，消息生存期分别为 3 h、6 h、12 h、18 h、24 h，消息的源和目标根据消息生存期选择，从生存期内至少发生过一次联系的节点中随机挑选。实验比较的指标包括消息投递成功率和传输开销。投递成功率是成功到达目标的消息数占系统创建的原始消息数的比例，传输开销指投递失败的消息数与投递成功消息数的比例。

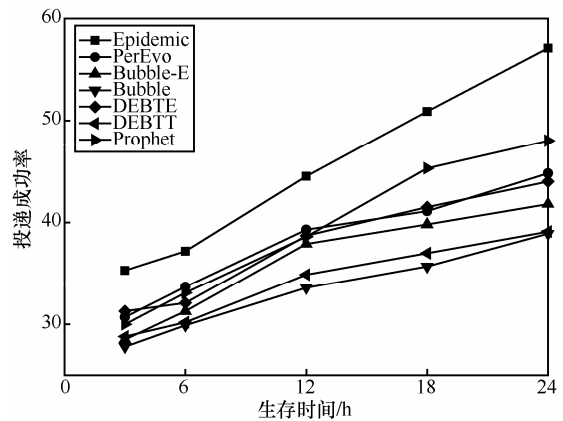
评估中多种路由策略参与了比较：Epidemic、Prophet、Bubble、DEBT 及 PerEvo。Bubble 的社区检测基于分布式 k 团算法，中心性基于 c -window，为了突出比较，除原 Bubble 外，还提供了另一个增强版本 Bubble-E，其社区检测使用 τ -window 社区。DEBT 也提供了 2 个版本，DEBTE 和 DEBTT^[20]，其中 DEBTE 在 DEBT 所检测的社区内泛洪消息，但可能产生路由循环，而 DEBTT 通过构建层次式信息转发树彻底避免了循环。Prophet 参数使用文献 [5] 中的默认参数，其他详细实验参数如表 3 所示。

表 3 实验中的主要参数设置

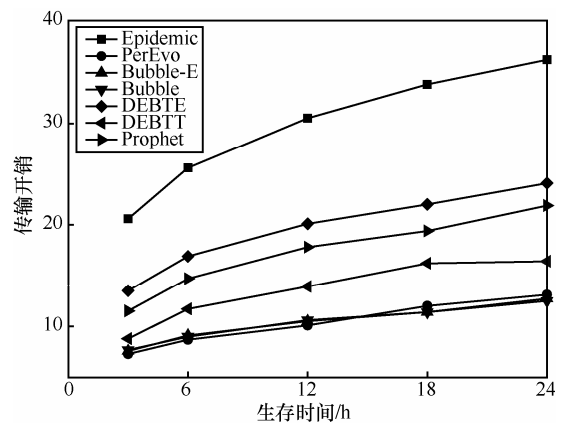
参数	取值
(PerEvo)时间窗上社区保存的周期数 L	5
(PerEvo)时间窗长度 τ	3 h
(PerEvo Bubble-E)社区老化指数 α	0.1
(PerEvo Bubble Bubble-E)社区阈值 F_{thresh}	300 s
(PerEvo)TTL 社区准入参数 λ	0.5
(DEBTT DEBTT)时间窗长度	1 h
(DEBTE DEBTT)基线计算时间窗数 w	6
(PerEvo Bubble Bubble-E) k 团 k 值	5

7.2 总体性能评估

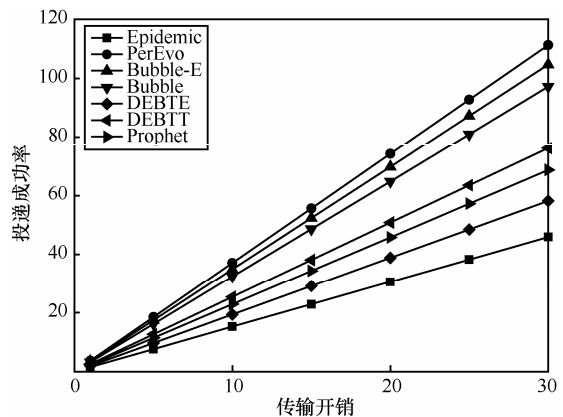
本节评估 PerEvo 与其他路由机制在整体性能上的比较，图 12(a)和图 12(b)是各协议下消息投递成功率和传输开销随消息生存期变化的情况，由图 12 可知，虽然 Prophet 在投递成功率上与 PerEvo 等社会特征路由差别不大，但在传输开销上远远大于社会特征路由。从图 12 还可看出，Bubble-E 在消息



(a) 投递成功率



(b) 传输开销



(c) 投递成功率与传输开销的关系

图 12 各种路由机制消息投递成功率和传输开销的比较

投递成功率上优于 Bubble，主要在于 Bubble-E 避免了超大社区问题，而 PerEvo 在消息投递成功率上优于 Bubble-E，主要源于 2 个方面：第一，TTL 社区较 τ -window 社区能够更准确预见节点在消息生存期上的社会关系；第二，TTL 中心性较 c -window 能更准确反映节点在消息生存期上的中继能力。

为了突出效果，图 12(c)比较了各策略消息投递的性能，即投递成功率与传输开销的关系，其中，投递成功率和传输开销分别采用不同生存期得到的实验结果的均值表示，从图中可以看出，PerEvo 是所有比较的路由协议中性能最优的。

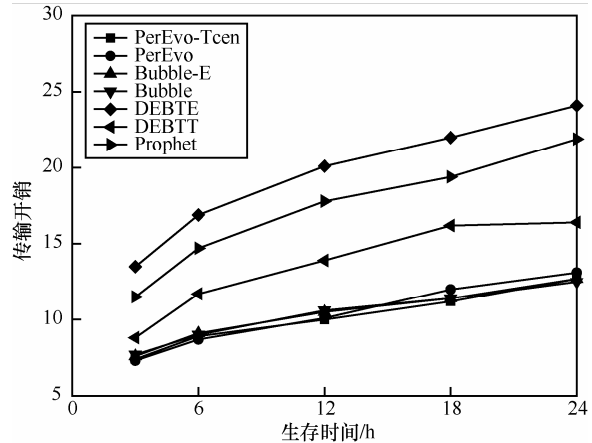
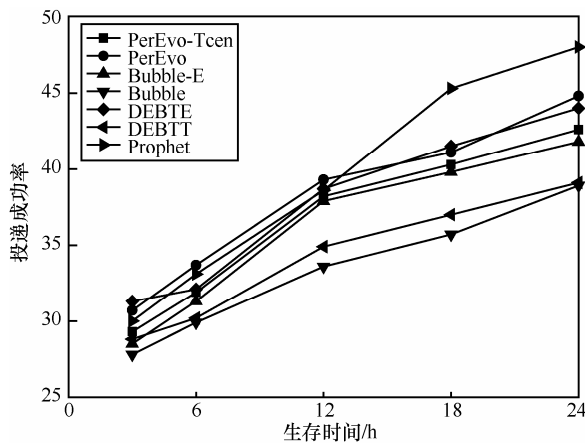
7.3 TTL 社区对路由影响的评估

本节评估 TTL 社区对路由性能的影响，为此，实现了 PerEvo 的缩减版 PerEvo-Tcen，PerEvo-Tcen 的中心性采用 TTL 中心性，社区采用 τ -window 社区检测方法，图 13 比较了该版本与 PerEvo、Bubble

及 Bubble-E 的消息投递成功率和传输开销。由图中可知，PerEvo-Tcen 传输开销与 PerEvo 相差不多，但投递成功率明显低于 PerEvo，这说明 TTL 社区较 τ -window 社区能够更好地预测节点在消息生存期内的社会联系，在保持传输开销变化不大的情况下，有效提高路由投递的成功率。

7.4 TTL 中心性对路由的影响评估

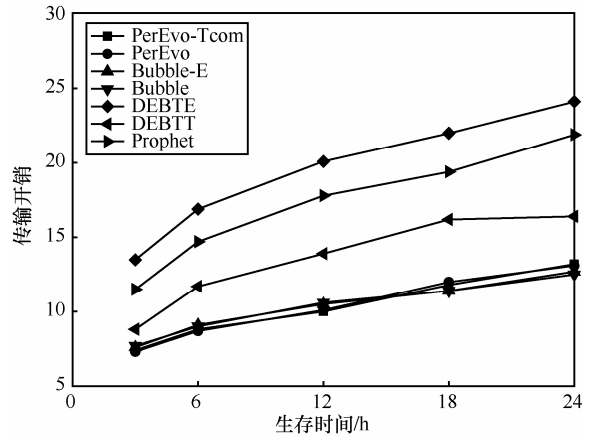
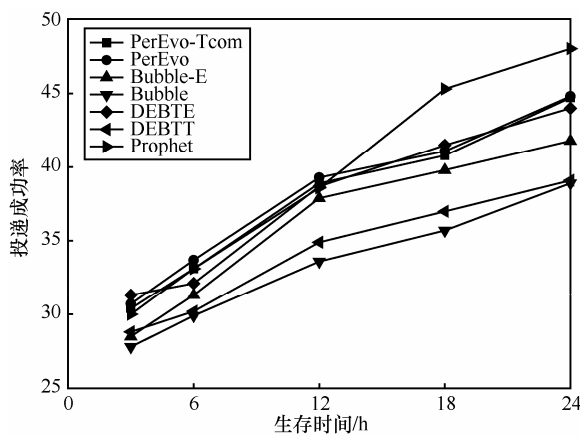
本节评估 TTL 中心性对路由性能的影响，为此，实现了 PerEvo 的一个缩减版 PerEvo-Tcom，PerEvo-Tcom 的社区采用 TTL 社区，中心性采用 c -window。图 14 比较了 PerEvo-Tcom 与 PerEvo、Bubble 及 Bubble-E 的消息投递成功率和传输开销，由图 14 可知，PerEvo-Tcom 在传输开销上与 PerEvo 相差不多，在投递成功率上要低于 PerEvo，这说明 TTL 中心性较 c -window 中心性更能够准确反映节点在消息生存期内的中继能力，在保持传输开销变化不大的情况下，提高了消息投递的成功率。



(a) 投递成功率

(b) 传输开销

图 13 TTL 社区对路由性能的影响评估



(a) 投递成功率

(b) 传输开销

图 14 TTL 中心性对路由性能的影响评估

8 结束语

基于节点社会特征周期演化特性,本文提出了改进消息转发性能的转发测度:TTL社区和TTL中心性,以更准确地预测节点在消息生存期内的社会联系。基于Trace数据集上的大量实验表明,与现有的基于社会特征的路由策略比较,基于TTL社区和TTL中心性的PerEvo路由转发策略在保证较低消息传输开销的情况下,能够有效提高机会移动网络消息投递的成功率。

分布式社区检测算法是机会移动网络路由的重要研究方向,社区中哪些因素对于路由性能起主要影响有待继续深入研究,特别是,如何有效建模节点在动态社区演化中的关键特性并从理论上分析这些特性与节点中继消息的能力之间的相关关系是本研究下一步的方向。

参考文献:

- [1] HUI P, CHAINTREAU A, SCOTT J, *et al.* Pocket switched networks and human mobility in conference environments[A]. Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking[C]. ACM: Philadelphia, Pennsylvania, USA, 2005.
- [2] MCNAMARA L, MASCOLO C, CAPRA L. Media sharing based on colocation prediction in urban transport[A]. Proceedings of the 14th ACM International Conference on Mobile Computing and Networking[C]. ACM: San Francisco, California, USA, 2008.58-69.
- [3] PELUSI L, PASSARELLA A, CONTI M. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks[J]. IEEE Communications Magazine, 2006, 44(11):134-141.
- [4] FALL K. A delay-tolerant network architecture for challenged internets[A]. Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications[C]. ACM: Karlsruhe, Germany, 2003.
- [5] LINDGREN A, DORIA A, SCHELEN O. Probabilistic routing in intermittently connected networks[J]. SIGMOBILE Mob Comput Commun Rev, 2003, 7(3): 19-20.
- [6] BURGESS J, GALLAGHER B, JENSEN D, *et al.* MaxProp: routing for vehicle-based disruption-tolerant networks[A]. Proceedings of INFOCOM[C]. Barcelona, Spain, 2006.
- [7] HUI P, CROWCROFT J, YONEKI E, BUBBLE Rap: social-based forwarding in delay-tolerant networks[J]. IEEE Transactions on Mobile Computing, 2011, 10(11): 1576-1589.
- [8] DALY E M, HAAHR M. Social network analysis for routing in disconnected delay-tolerant MANET[A]. Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing[C]. ACM: Montreal, Quebec, Canada, 2007.
- [9] NEWMAN M E, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004,69(2):026113.
- [10] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social Networks, 1979, 1(3): 215-239.
- [11] HUI P, CROWCROFT J. How small labels create big improvements[A]. Proceedings of IEEE PerCom Workshops[C]. New York, USA, 2007.
- [12] SONG C M, QU Z, BLUMM N, *et al.* Limits of predictability in human mobility[J]. Science, 2010. 327(5968): 1018-1021.
- [13] EAGLE N, PENTLAND A. Reality mining: sensing complex social systems[J]. Personal Ubiquitous Comput, 2006, 10(4):255-268.
- [14] MOREIRA W, MENDES P, SARGENTO S. Opportunistic routing based on daily routines[A]. Proceedings of IEEE WoWMoM[C]. San Francisco, California, USA, 2012.
- [15] FREEMAN L C. A set of measures of centrality based on betweenness[A]. Sociometry[C]. 1977.35-41.
- [16] MARSDEN P V. Egocentric and sociocentric measures of network centrality[J]. Social Networks, 2002. 24(4): 407-422.
- [17] HUI P, YONEKI E, CHAN S Y, *et al.* Distributed community detection in delay tolerant networks[A]. Proceedings of 2nd ACM/IEEE International Workshop on Mobility in the Evolving Internet Architecture[C]. Kyoto, Japan, 2007.
- [18] PALLA G, DERENYI I, FARKAS I, *et al.* Uncovering the overlapping community structure of complex networks in nature and society[A]. Proceeding of The 2nd ACM/IEEE Internet Workshop on Mobility in the Evolving Internet Architecture[C]. Kyoto, Japan, 2007.
- [19] ORLINSKI M, FILER N. Quality distributed community formation for data delivery in pocket switched networks[A]. Proceedings of the Fourth Annual Workshop on Simplifying Complex Networks for Practitioners[C]. Lyon, France, 2012.
- [20] ORLINSKI M, FILER N. Distributed expectation-based spatio-temporal cluster detection for pocket switched networks[A]. Proceedings of Wireless Days IFIP IEEE[C]. 2012.
- [21] GAO W, LI Q, ZHAO B, *et al.* Social-aware multicast in disruption-tolerant networks[J]. IEEE/ACM Transactions on Networking, 2012, 20(5):1553-1566.
- [22] GAO W, CAO G T L P, *et al.* On exploiting transient social contact patterns for data forwarding in delay-tolerant networks[J]. IEEE Transactions on Mobile Computing, 2013, 12(1):151-165.
- [23] KIM H, ANDERSON R. Temporal node centrality in complex networks[J]. Physical Review E, 2012. 85(2): 026107.
- [24] KERANEN A, OTT J, KARKKAINEN T. The ONE simulator for DTN protocol evaluation[A]. Proceedings of the 2nd International Conference on Simulation Tools and Techniques[C]. Rome, Italy, 2009.

作者简介:



黄永锋 (1978-), 男, 江苏溧阳人, 东南大学博士生, 主要研究方向为云计算环境下的内容分发、时间演化的动态连通网络等。

董永强 [通信作者] (1973-), 男, 河南浉池人, 博士, 东南大学副研究员, 主要研究方向为网络体系结构、移动网络计算。E-mail: dongyq@eu.edu.cn。

张三峰 (1979-), 男, 山东金乡人, 博士, 东南大学讲师, 主要研究方向为 P2P 网络、信任管理、移动自组网等。

吴国新 (1956-), 男, 安徽芜湖人, 东南大学教授、博士生导师, 主要研究方向为网络协议、网络安全、自组网等。